# Whole Genome Sequencing: Innovation Dream or Privacy Nightmare?

**Erman Ayday[1], Emiliano De Cristofaro[2], Jean-Pierre Hubaux[1], Gene Tsudik[3]**

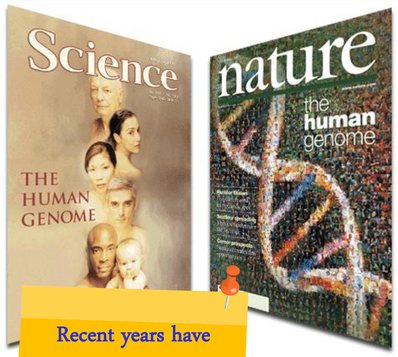
[1]{firstname.lastname@epfl.ch}, [2]me@emilianodc.com, [3]gts@ics.uci.edu

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

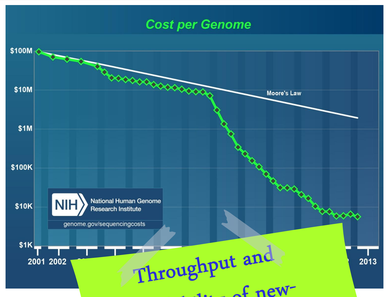
parc A Xerox Company

UCIrvine University of California, Irvine

## Introduction and Background



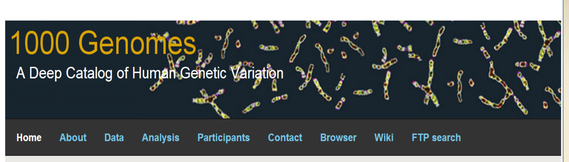

Recent years have witnessed impressive advances in DNA sequencing




Throughput and affordability of new-generation sequencing platforms have increased at a pace faster than Moore's Law

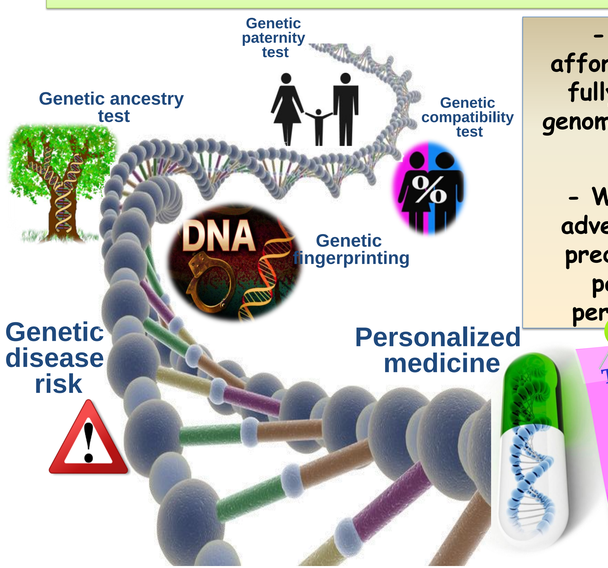
ENCODE

Personal Genome Project


- More research is needed to understand the complex relationship between genome and health
- Scientific community needs large cohorts of patients willing to share their genetic material
- All participants must agree to have their genomic data made publicly available on the Internet

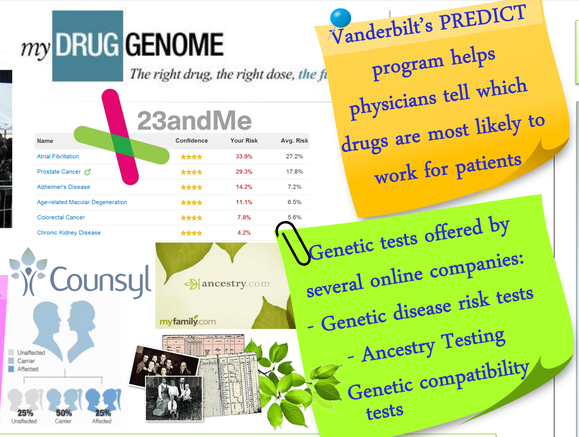
## Innovation Dream?




- Widespread and affordable availability of fully sequenced human genomes creates enormous opportunities
- WGS facilitates the advent of a new era of predictive, preventive, participatory, and personalized medicine

Treatment and medication type/dosage can be tailored to the precise genetic makeup of patients


BRCA1/BRCA2 genes are in correlation to familial breast and ovarian cancer syndromes

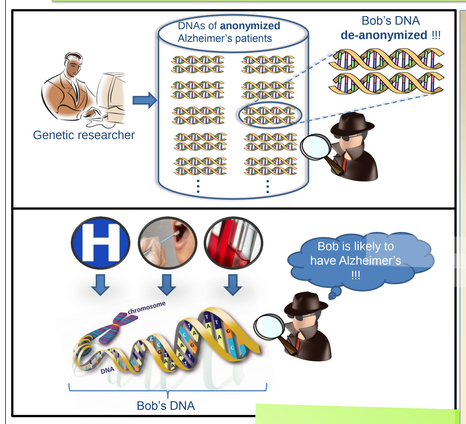



Vanderbilt's PREDICT program helps physicians tell which drugs are most likely to work for patients

Genetic tests offered by several online companies:
- Genetic disease risk tests
- Ancestry Testing

Genetic compatibility tests

## Privacy Nightmare?



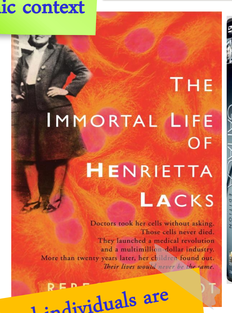

- DNA reflects information about genetic predispositions
- DNA contains information about ancestors and progeny
- DNA does not change over time, hence revoking or replacing it is impossible
- DNA analysis used in law enforcement and healthcare prompts ethical issues
- It is hard to estimate the extent of the personal information that could be extracted from the genome in the future

De-identification or aggregation are ineffective in the genomic context

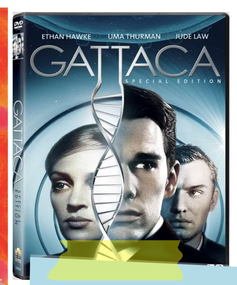
**Identifying Personal Genomes by Surname Inference**

Melissa Gymrek,[1,2,3,4] Amy L. McGuire,[5] David Golan,[6] Eran Halperin,[7,8,9] Yaniv Erlich[3+]

Re-identifying DNA donors from a public research database using popular genealogy Web sites and other information



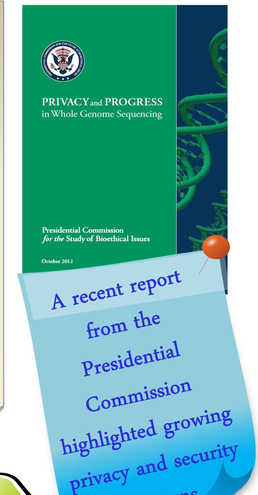

- Genomes of any two closely related individuals are more that 99.9% identical
- Controversy between the family of deceased Henrietta Lacks (whose genome was sequenced and published after her death) and the scientists who are in favor of publishing genomes online for research purposes




1997 movie denotes the theory that distinctive human characteristics and abilities are determined by genes, resulting in genetic discrimination

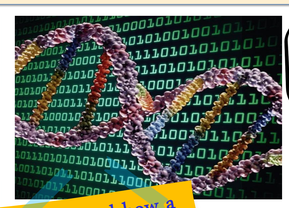
## Open Research Questions

- Storage and Accessibility: Genome at Rest
  - Personal device vs. cloud
  - Clear vs. encrypted
  - Backups and copies of genomes
  - Integrity and authenticity
- Privacy: Genome in Action
  - Privacy vs. utility
  - Cryptographic algorithms for genetic tests
  - Kinship and consequences
  - Flexibility for future discoveries
- Long-term data protection
  - Non-cryptographic approaches
  - Re-encryption and erase of a genome
- Accuracy and Accountability
- Efficiency
- Usability
- Large-scale research on human genomes




A recent report from the Presidential Commission highlighted growing privacy and security concerns

Even the best of the cryptographic algorithms we use today could be broken in around 30 years



Where and how a digitalized genome should be stored?

USABILITY



A privacy-preserving paternity test on digitized genomes should be accurate and admissible in a court of law

## Next Steps

Mitigating privacy issues will require long-term collaboration among geneticists, other healthcare providers, ethicists, lawmakers, and computer scientists


New funding programs need to target this topic to foster the collaboration

- If not addressed, privacy issues addressed here could affect numerous individuals
- By impeding genomic research, it would also affect the well-being of our society as a whole
- There is an urgent need for collaboration among researchers in various fields

For more information:
http://lca.epfl.ch/projects/genomic-privacy/
http://emiliano.im/
http://www.ics.uci.edu/~gts/